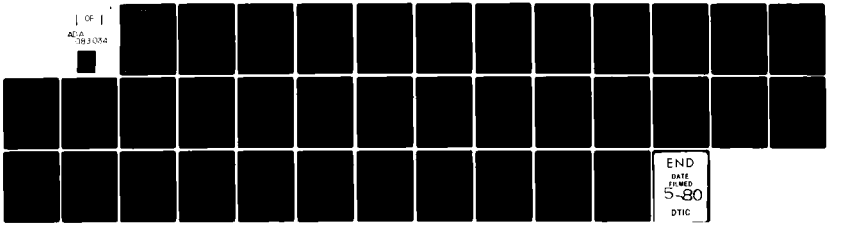
AD-A083 034 OHIO STATE UNIV COLUMBUS DEPT OF GEODETIC SCIENCE F/G 8/5
OPTIMAL ESTIMATION FROM DATA REGULARLY SAMPLED ON A SPHERE WITH--ETC(U)
SEP 79 O L COLOMBO F19628-79-C-0027
UNCLASSIFIED DGS-291 AFGL -TR-79-0227 NL



AFGL-TR-79-0227

LEVEL II

12
B.S.

OPTIMAL ESTIMATION FROM DATA REGULARLY SAMPLED ON A SPHERE WITH APPLICATIONS IN GEODESY

Oscar L. Colombo

The Ohio State University
Research Foundation
Columbus, Ohio 43212

September, 1979

Scientific Report No. 1

AIR FORCE GEOPHYSICS LABORATORY
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
HANSCOM AFB, MASSACHUSETTS 01731

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

(19) REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
(18) 1. REPORT NUMBER AFGL-TR-79-0227 ✓	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
(6) 4. TITLE (and Subtitle) OPTIMAL ESTIMATION FROM DATA REGULARLY SAMPLED ON A SPHERE WITH APPLICATIONS IN GEODESY		5. TYPE OF REPORT & PERIOD COVERED Scientific. Interim Scientific Report No. 1
		6. PERFORMING ORG. REPORT NUMBER Dept. of Geod. Sci. No. 291
(10) 7. AUTHOR(s) Oscar L. Colombo		(15) 8. CONTRACT OR GRANT NUMBER(s) F19628-79-C-0027
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Geodetic Science The Ohio State University - 1958 Neil Avenue Columbus, Ohio 43210		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F (16) 2309G1AW (17) G1
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Geophysics Laboratory Hanscom AFB, Massachusetts 01731 Contract Monitor: Bela Szabo/LW		(11) 12. REPORT DATE Sep[redacted]79
		13. NUMBER OF PAGES 29 pages
(9) 14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	(12) 36	15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/ DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

14 DG-211, SCIENTIFIC-1

A - Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

geodesy, estimation theory, gravity anomalies, potential coefficients

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The size of the variance-covariance matrix of the data, used to obtain minimum variance estimators for collocation, is as large as the number of observations in the data set. For some arrangements of the data, such as the usual "equal angle" (or "regular") grid, the matrix presents a very strong Toeplitz-circulant structure that can be exploited to reduce computing in setting-up and inverting the matrix. This reduction can be quite drastic. This report discusses such structure and presents

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

an algorithm for implementing collocation efficiently. Three applications are considered: (a) the spherical harmonic analysis of point data; (b) the same analysis using area means; (c) the estimate of the disturbing potential from gravity anomalies. The harmonic analysis is optimal for noisy data as well; with noiseless data it provides harmonic coefficients with minimum aliasing.

Foreword

This report was prepared by Dr. Oscar Colombo, Post Doctoral Researcher, Department of Geodetic Science, The Ohio State University, under Air Force Contract No. F19628-79-C-0027, The Ohio State University Research Foundation Project No. 711664, Project Supervisor Richard H. Rapp. The contract covering this research is administered by the Air Force Geophysics Laboratory, Hanscom Air Force Base, Massachusetts, with Mr. Bela Szabo, Contract Monitor.

Acknowledgements

The author is thankful to Professor Richard H. Rapp and to Dr. Reiner Rummel and Dr. Hans Sünkel for their many comments and the encouragement they gave to this research. Pamela Pozderac is thanked for her typing of the originals, with their less than optimal notation.

Table of Contents

1. Introduction

The minimum variance linear estimates of values of a variable $s(\varphi,\lambda,r)$ from a finite set of measurements consisting of samples of a signal $z(\varphi,\lambda,r)$ plus measurement noise $n(\varphi,\lambda,r)$, can be obtained using the following formulas (Moritz, 1972):

$$\begin{aligned} \hat{\underline{s}} &= F\underline{d} \\ F &= C_{sz}\,C_{dd}^{-1} \\ C_{dd} &= C_{zz} + D \end{aligned} \tag{1.1}$$

The variance-covariance matrix of the estimation errors is

$$\begin{aligned} E &= M\{(\hat{\underline{s}} - F\underline{d})(\hat{\underline{s}} - F\underline{d})^T\} \\ &= C_{ss} - FC_{sz}^T - C_{sz}F^T + FC_{dd}F^T \end{aligned} \tag{1.2-a}$$

or

$$E = C_{ss} - C_{sz}C_{dd}^{-1}C_{sz}^T \tag{1.2-b}$$

in the case of the optimal estimator F as given in (1.1) above. Here

$M\{\ \}$	is some kind of average over the sphere,
$\hat{\underline{s}}$	is the N_s vector of estimates,
$\underline{d}$	is the N_d data vector,
C_{sz}	is the $N_s \times N_d$ covariance matrix of true values of s and z,
C_{dd}	is the $N_d \times N_d$ data covariance matrix,
C_{zz}	is the $N_d \times N_d$ measurements' signal covariance matrix,
D	is the $N_d \times N_d$ measurements' noise variance-covariance matrix,
C_{ss}	is the $N_s \times N_s$ covariance matrix of the true values of s.

All variables are supposed to have zero mean values: $M\{s\} = M\{z\} = M\{n\} = M\{\hat{\underline{s}}\} = 0$, while the noise is assumed to be uncorrelated both with the signal $s(\varphi,\lambda,r)$ and with $z(\varphi,\lambda,r)$. The estimates $\hat{\underline{s}}$ obtained by using the estimator F defined in (1.1) have a corresponding E matrix whose diagonal elements are the smallest for all possible linear estimators with the same data pattern. In this sense, F is optimal. This estimator depends on the particular $M\{\ \}$ chosen, as explained by Rummel and Schwarz (1977), because this affects the elements of the covariance matrices. In turn, these influence the actual estimates $\hat{\underline{s}}$, and E as well.

A major problem with this method, also known as collocation, is that the number of measurements N_d is also the dimension of C_{dd}, so setting up this matrix requires computing up to $\frac{1}{2}N_d^2$ different elements. These elements are given by the expression

$$\begin{aligned} c_{dd}^{ij} &= M\{z(\varphi_i,\lambda_i,r_i)\,z(\varphi_j,\lambda_j,r_j)\} + M\{n(\varphi_i,\lambda_i,r_i)\,n(\varphi_j,\lambda_j,r_j)\} \\ &= c_{zz}(P_i,P_j) + c_{nn}(P_i,P_j) \end{aligned} \tag{1.3}$$

where c_{zz} is the "covariance function" of the signal z, and c_{nn} is that of the

noise n. Both depend, generally, on the sampling points P_i and P_j. In most cases

$$c_{nn}(P_i, P_j) = \begin{cases} \sigma_i & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \tag{1.4}$$

which is what is meant by the words "white noise". Calculating the covariance functions' values may involve several operations in a computer. Sometimes a large number of terms of a series expansion may have to be evaluated; even with closed expressions this can be a time consuming enterprise. Moreover, solving the "normal" equations

$$C_{dd}\, F^T = C_{sz}^T \tag{1.5}$$

to find F requires an additional $k N^3$ operations, k being a constant characteristic of the method used. Both tasks can be formidable with large numbers of measurements, presenting the paradox that, while more data must result (theoretically) in better estimates, these are harder to obtain and are worse affected by numerical errors. The discussion that follows will show how, for certain types of covariances and certain distributions of data, the problem becomes manageable even with large data sets.

2. Limitations on the Data Arrangement and on the Covariance Functions

Geodetic data, such as gravity anomalies, geoid undulations, etc., are given usually in the form of either point values or of area means. In each case let the following conditions be satisfied:

Point Values	Area Means
C-1 All data points are on the same sphere of radius R;	C-1' All area averages are taken over blocks on the same sphere of radius R;
C-2 All data points are nodes in a grid of "parallels and meridians" (Fig. 2.1) with poles excluded;	C-2' All blocks belong to a "parallels and meridians" partition of the sphere without circular blocks (polar caps) about the poles;
C-3 No data point in a row (all nodes along the same parallel) is empty;	C-3' In a row of blocks (blocks bound by the same parallels) if there is data in one, there is data in all;
C-4 The longitude increment between adjacent meridians is constant.	C-4' All blocks have the same longitude span.

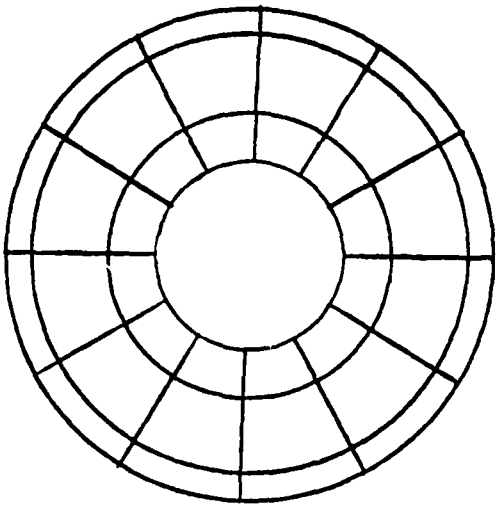


Figure 2.1-a. Example of a grid of point measurements seen from one of the poles. Notice that the Pole itself is excluded, and that the separation of the "meridians" is constant, not so that of the "parallels".

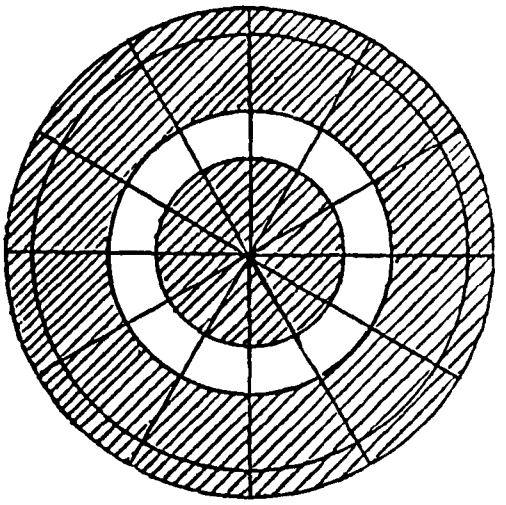


Figure 2.1-b. Grid of area means. The shaded blocks contain data. The "parallels" and "meridians" delimiting the blocks are the same as in Figure 1.1-a. Notice the absence of an undivided "polar cap".

In what follows, data points along the same parallel (blocks between the same bounding parallels) form a row, while those along (between) the same (bounding) meridian(s) form a column. Rows are numbered from N to S, and columns from W to E. The latitude increments between data points (span of the blocks in latitude) do not have to be constant, but the separation in longitude (longitude span of the blocks) has to be constant. Moreover, rows of blocks with data can be separated by empty ones.

All rows must have the same number of points (blocks), equal to the number of columns N_c, which is also the number of meridians. A meridian is a 180° arc from pole to pole. Calling the number of rows containing data N_r, then the total number of non-empty points (blocks) in the grid is

$$N_d = N_c \times N_r$$

This is the dimension of $\underline{d}$ and also that of C_{dd}, or the number of values in the data set.

Having explained the kind of grid admissible,[1] something must be said about the covariance functions (or the $M\{\ \}$ operator) that can be used. They must satisfy the following restrictions:

[1] In general, the data can be irregularly distributed, but point values can be estimated at the nodes of the grid from neighboring measurements, and area means from the averages of measurements on the same blocks. The estimates' variances should be found, also, to set up D.

Point Values	Area Means
C-5 Given two rows i and j (including i = j), the value of the covariance function $c_{xx}(P_{ik}, P_{jm}) = M\{x_{ik}\, x_{jm}\}$ between points P_{ik} and P_{jm} must depend only on $\lvert\lambda_k - \lambda_m\rvert$ (including k = m).	C-5' Given two rows of blocks, i and j (including i = j), the covariance between two area means $\bar{x}_{ik}$ and $\bar{x}_{jm}$ must depend only on $\lvert\lambda_{ik} - \lambda_{jm}\rvert$ (including k=m), where λ_{ik} is the longitude of the W boundary of block ik .

Restriction C-5 (C-5') allows a variety of covariance functions, including the so-called "isotropic" or "global" used by the majority of workers at present (Rummel and Schwarz, 1977).

Usually the "noise" function n is supposed to be uncorrelated from measurement to measurement ("white noise") resulting in a diagonal D matrix. To satisfy C-5 (or C-5') the variance of the errors must be the same at all points (blocks) on the same row, while it can vary from row to row. This assumption is rather restrictive, as in practice (particularly with area means, where the size of the block changes with latitude) the standard deviation of the measurements can vary from place to place. Nearly homogeneous data sets may become more common through an increase in the measuring of the gravity field from satellites (satellite altimetry, satellite-satellite tracking, etc.). In some cases, even when C-5 or C-5' are not exactly fulfilled, the noise fluctuations along rows might be small, and the mean standard deviation of each row could be used to set up D. A refinement of this idea is explained in the Appendix.

3. The Structure of the Data Covariance Matrix

When all the limitations described in the previous section are present, matrices C_{zz}, D and $C_{dd} = C_{zz} + D$ all have the same well-defined structure. To make it clear, let us arrange the measurements in d as follows[1]:

$$\underline{d} = [\underline{d}_1\ \underline{d}_2 \ldots \underline{d}_i \ldots \underline{d}_{N_r}]^T$$

where

$$\underline{d}_i = [d_{i0}\, d_{i1} \ldots d_{ik} \ldots d_{iN_c-1}]^T$$

so the $\underline{d}_i$ are N_r subvectors of dimension N_c, each containing the data values for one row. This brings about the corresponding partitioning of C_{dd}, C_{zz} and D into N_r^2 $N_c \times N_c$ "row submatrices" C_{ij}, containing all the correlations between points (blocks) in the i th and the j th row. Assume N = 5, that the point on the "0 th meridian" is the first point in any row, and that all others in the same row are ordered, like their meridians, clockwise when seen from the

[1] To simplify typing $[\underline{d}_1 \ldots \underline{d}_{N_r}]^T$ is used throughout this work instead of $[\underline{d}_1^T \ldots \underline{d}_{N_r}^T]^T = \begin{bmatrix} \underline{d}_1 \\ \vdots \\ \underline{d}_{N_r} \end{bmatrix}$

North Pole (i.e., along increasing longitudes). Then, the point immediately to the West of the first is the last, or N_c th, in any row. From the various restrictions mentioned in the previous section, both for points and blocks, it follows that if

$$\begin{aligned} a &= c_{xx}(x_{i0}, x_{j0}) \\ b &= c_{xx}(x_{i0}, x_{j1}) = c_{xx}(x_{i0}, x_{j4}) \\ c &= c_{xx}(x_{i0}, x_{j2}) = c_{xx}(x_{i0}, x_{j3}) \end{aligned}$$

are the covariances between the first point (block) in row i and all the points (blocks) in row j, then

$$C_{ij} = \begin{bmatrix} a & b & c & c & b \\ b & a & b & c & c \\ c & b & a & b & c \\ c & c & b & a & b \\ b & c & c & b & a \end{bmatrix}$$

Calling p the row subscript and q the column subscript of the c_{pq}^{ij} elements of this matrix, we notice that they have the following properties:

$$\begin{aligned} &c_{pq}^{ij} = c_{p+1\,q+1}^{ij}; \quad c_{p1}^{ij} = c_{p-1\,N_c}^{ij} \quad \text{when } p > 1; \\ &c_{pq}^{ij} = c_{p\,N_c+2-q} \quad \text{when } p = 1,\ q > 1 \end{aligned}$$

Matrices of this type belong to the class known as "circular" or "Toeplitz circulant" (Lancaster, 1969). As C_{dd}, C_{zz} and D consist of blocks of this kind, they can be described as block matrices of Toeplitz circulant blocks. As shown in the remainder of this work, these matrices are much easier to set up and to invert than ordinary symmetric matrices.

The first row in C_{ij} resembles a succession of equispaced samples of some even function, so it can be represented exactly using a finite sum of cosines:

$$c_{1q}^{ij} = \sum_{k=0}^{N} a_k^{ij} \cos \frac{2\pi k}{N_c} q \qquad (3.1)$$

where $2N = N_c$ if N_c is even; $2N + 1 = N_c$ if N_c is odd, and

$$a_k^{ij} = \frac{1}{H} \sum_{q=0}^{N_c-1} c_{1q}^{ij} \cos \frac{2\pi k}{N_c} q, \text{ where H is defined later.} \qquad (3.2)$$

The pth row is the same as the first rotated p times to the right:

$$\begin{aligned} c_{pq}^{ij} &= \sum_{k=0}^{N} a_k^{ij} \cos \frac{2\pi k}{N_c} (q-p) \\ &= \sum_{k=0}^{N} a_k^{ij} \cos \frac{2\pi k}{N_c} q \cos \frac{2\pi k}{N_c} p + \sum_{k=0}^{N} a_k^{ij} \sin \frac{2\pi k}{N_c} q \sin \frac{2\pi k}{N_c} p \end{aligned} \qquad (3.3)$$

From the well-known trigonometric expressions

$$\sum_{q=0}^{N_c-1} \begin{pmatrix}\cos\\ \sin\end{pmatrix} \frac{2\pi\alpha}{N_c} q \begin{pmatrix}\cos\\ \sin\end{pmatrix} \frac{2\pi k}{N_c} q = H = \begin{cases} N_c & \text{if } \alpha = k = 0 \text{ or if } \alpha = k = N \ (N_c \text{ even}) \\ N_c/2 & \text{if } 0 < \alpha = k < N \ (=N \text{ if } N_c \text{ odd}) \\ 0 & \text{if } 0 \le \alpha \ne k \le N \end{cases}$$

$$\sum_{q=0}^{N_c-1} \cos \frac{2\pi\alpha}{N_c} q \sin \frac{2\pi k}{N_c} q = 0 \quad \text{always}$$

follows that

$$a_k^{ij} H \cos \frac{2\pi k}{N_c} p = \sum_{q=0}^{N_c-1} c_{pq}^{ij} \cos \frac{2\pi k}{N_c} q \tag{3.4-a}$$

and

$$a_k^{ij} H \sin \frac{2\pi k}{N_c} p = \sum_{q=0}^{N_c-1} c_{pq}^{ij} \sin \frac{2\pi k}{N_c} q \tag{3.4-b}$$

for $0 \le k \le N$. Consider now the following vectors:

$$\underline{c}_k = [1, \cos \frac{2\pi k}{N_c}, \cos \frac{4\pi k}{N_c}, \ldots, \cos \frac{2\pi k}{N_c} q, \ldots, \cos \frac{2\pi k}{N_c}(N_c-1)]^T$$

and

$$\underline{s}_k = [0, \sin \frac{2\pi k}{N_c}, \sin \frac{4\pi k}{N_c}, \ldots, \sin \frac{2\pi k}{N_c} q, \ldots, \sin \frac{2\pi k}{N_c}(N_c-1)]^T$$

Expressions (3.4-a) and (3.4-b) can be written in matrix form

$$a_k^{ij} H \underline{c}_k = C_{ij} \underline{c}_k \tag{3.5-a}$$

$$a_k^{ij} H \underline{s}_k = C_{ij} \underline{s}_k \tag{3.5-b}$$

Consequently, vectors such as $\underline{c}_k$ and $\underline{s}_k$ are <u>eigenvectors</u> of the submatrix C_{ij}, and to each such pair ($k = 0, 1, \ldots, N$) corresponds <u>one eigenvalue</u>

$$\lambda_k^{ij} = a_k^{ij} H \tag{3.6}$$

3.1 A Fundamental Property

Consider the N_d-vector

$$\underline{f}_\alpha^k = [\underline{f}_{\alpha,1}^k \underline{f}_{\alpha,2}^k \cdots \underline{f}_{\alpha,i}^k \cdots \underline{f}_{\alpha,N_r}^k]^T \tag{3.7}$$

where the i th partition is $\underline{f}_{\alpha,i}^k = \varphi_i^k \underline{c}_k$ if $\alpha = 0$, or $\underline{f}_{\alpha,i}^k = \varphi_i^k \underline{s}_k$ if $\alpha = 1$. The product $\underline{g}_\alpha^k = C_{dd} \underline{f}_\alpha^k$ can be partitioned in the same way:

$$\underline{g}_\alpha^k = [\underline{g}_{\alpha,1}^k \underline{g}_{\alpha,2}^k \cdots \underline{g}_{\alpha,i}^k \cdots \underline{g}_{\alpha,N_r}^k]^T \tag{3.8}$$

where the i th partition is

$$\underline{g}_{\alpha i}^k = \sum_{j=1}^{N_r} C_{ij} \underline{f}_{\alpha,j}^k = \sum_{j=1}^{N_r} \varphi_j^k C_{ij} \begin{pmatrix}\underline{c}_k\\ \underline{s}_k\end{pmatrix} = \left(\sum_{j=1}^{N_r} a_k^{ij} H \varphi_j^k\right) \begin{pmatrix}\underline{c}_k\\ \underline{s}_k\end{pmatrix} = \gamma_i^k \begin{pmatrix}\underline{c}_k\\ \underline{s}_k\end{pmatrix} \tag{3.9-a}$$

with $$\gamma_i^k = \sum_{j=1}^{N_r} a_k^{ij} H \varphi_j = \sum_{j=1}^{N_r} \lambda_k^{ij} \varphi_j^k \tag{3.9-b}$$

For a given α and k, $\underline{f}_\alpha^k$ is in a one-to-one relationship with the N_r-vector $\underline{\varphi}(k)$:

$$\underline{\varphi}(k) = [\varphi_1^k\ \varphi_2^k \ldots \varphi_i^k \ldots \varphi_{N_r}^k]^T \Longleftrightarrow \underline{f}_\alpha^k = \left[\varphi_1^k\left(\frac{c_k}{s_k}\right)\ldots\varphi_{N_r}^k\left(\frac{c_k}{s_k}\right)\right]^T \qquad (3.10\text{-a})$$

and similarly

$$\underline{\gamma}(k) = [\gamma_1^k\ \gamma_2^k \ldots \gamma_i^k \ldots \gamma_{N_r}^k]^T \Longleftrightarrow \underline{g}_\alpha^k = C_{dd}\,\underline{f}_\alpha^k = \left[\gamma_1^k\left(\frac{c_k}{s_k}\right)\ldots\gamma_{N_r}^k\left(\frac{c_k}{s_k}\right)\right]^T \qquad (3.10\text{-b})$$

Expression (3.9) can be given matrix form:

$$\underline{\gamma}(k) = R(k)\underline{\varphi}(k) \Longleftrightarrow C_{dd}\,\underline{f}_\alpha^k \qquad (3.11)$$

where $R(k)$ is a $N_r \times N_r$ matrix with elements

$$r_{ij}^k = a_k^{ij} H = \lambda_k^{ij} \qquad (3.12)$$

Expressions (3.10-a), (3.10-b), (3.11) and (3.12) describe a property of the covariance matrix that is basic to the algorithm developed in the next two sections.

3.2 The Equation $\underline{y} = C_{dd}\,\underline{x}$

Consider once more the vectors $\underline{f}_\alpha^k$ and $\underline{g}_\alpha^k$ presented in the previous section, and the equation

$$\underline{g}_\alpha^k = C_{dd}\,\underline{f}_\alpha^k \qquad (3.13)$$

where the components of $\underline{f}_\alpha^k$ are the unknowns. As already explained, $\underline{f}_\alpha^k$ and $\underline{g}_\alpha^k$ are in one-to-one relationship to $\underline{\varphi}(k)$ and $\underline{\gamma}(k)$, respectively, so (3.13) can be regarded as equivalent to (3.11), because once we know

$$[\varphi_1^k\ \varphi_2^k \ldots \varphi_i^k \ldots \varphi_{N_r}^k]^T = \underline{\varphi}(k) = R(k)^{-1}\underline{\gamma}(k)$$

we know $\quad [\varphi_1^k\left(\frac{c_k}{s_k}\right)\varphi_2^k\left(\frac{c_k}{s_k}\right)\ldots\varphi_i^k\left(\frac{c_k}{s_k}\right)\ldots\varphi_{N_r}^k\left(\frac{c_k}{s_k}\right)]^T = \underline{f}_\alpha^k$

the solution to (3.13). If C_{dd}^{-1} exists, there is always a solution to (3.13), and thus to (3.11). Therefore, if C_{dd}^{-1} exists, so does $R(k)^{-1}$ for all $0 \le k \le N$. While the strict invertibility of the $R(k)$ is not essential, as long as $\underline{g}_\alpha^k$ is in the range of C_{dd}, the existence of inverses simplifies the argument. Assuming that C_{dd}^{-1} exists and that $\underline{y}$ is an arbitrary $N_c N_r$-vector, then there is another $N_c N_r$-vector $\underline{x}$ such that

$$\underline{y} = C_{dd}\,\underline{x} \qquad (3.14)$$

With the usual partition by data "rows":

$$\underline{x} = [\underline{x}_1\, \underline{x}_2 \cdots \underline{x}_i \cdots \underline{x}_{N_r}]^T \qquad \underline{y} = [\underline{y}_1\, \underline{y}_2 \cdots \underline{y}_i \cdots \underline{y}_{N_r}]^T$$

$$\underline{x}_i = [x_{i0}\, x_{i1} \cdots x_{ij} \cdots x_{iN_e-1}]^T \qquad \underline{y}_i = [y_{i0}\, y_{i1} \cdots y_{ij} \cdots y_{iN_e-1}]^T$$

A sequence of N_r numbers, such as the elements of $\underline{x}_i$ or $\underline{y}_i$, can be represented exactly by a sum of sine and cosine terms:

$$\underline{x}_{iq} = \sum_{k=0}^{N} c_{ik} \cos\frac{2\pi k}{N_r} q + \sum_{k=1}^{N} s_{ik} \sin\frac{2\pi k}{N_r} q \tag{3.15}$$

where N is as in (3.1), and $c_{iN} = 0$ if N_e is even. In matrix form:

$$x_{iq} = \sum_{k=0}^{N} c_{ik}\, \underline{c}_k + \sum_{k=1}^{N} s_{ik}\, \underline{s}_k \tag{3.16-a}$$

similarly

$$y_{iq} = \sum_{k=0}^{N} m_{ik}\, \underline{c}_k + \sum_{k=1}^{N} n_{ik}\, \underline{s}_k \tag{3.16-b}$$

Consequently, vectors such as $\underline{x}$ and $\underline{y}$ can be represented by sums of vectors of the same form as $\underline{f}_\alpha^k$ or $\underline{g}_\alpha^k$ above:

$$\underline{x} = \sum_{k=0}^{N} \sum_{\alpha=0}^{1} \underline{x}_\alpha^k \tag{3.17-a}$$

$$\underline{y} = \sum_{k=0}^{N} \sum_{\alpha=0}^{1} \underline{y}_\alpha^k, \quad (\underline{x}_0^N = \underline{y}_0^N = \underline{0} \text{ if } N_e \text{ is odd; } \underline{x}_1^0 = \underline{y}_1^0 = \underline{0} \text{ always}) \tag{3.17-b}$$

where

$$\underline{x}_\alpha^k = \left[\begin{pmatrix} c_{0k} & \underline{c}_k \\ 0 & \underline{s}_k \end{pmatrix} \begin{pmatrix} c_{1k} & \underline{c}_k \\ s_{1k} & \underline{s}_k \end{pmatrix} \cdots \begin{pmatrix} c_{ik} & \underline{c}_k \\ s_{ik} & \underline{s}_k \end{pmatrix} \cdots \begin{pmatrix} c_{N_r k} & \underline{c}_k \\ s_{N_r k} & \underline{s}_k \end{pmatrix} \right]^T$$

and

$$\underline{y}_\alpha^k = \left[\begin{pmatrix} m_{0k} & \underline{c}_k \\ 0 & \underline{s}_k \end{pmatrix} \begin{pmatrix} m_{1k} & \underline{c}_k \\ n_{1k} & \underline{s}_k \end{pmatrix} \cdots \begin{pmatrix} m_{ik} & \underline{c}_k \\ n_{ik} & \underline{s}_k \end{pmatrix} \cdots \begin{pmatrix} m_{N_r k} & \underline{c}_k \\ n_{N_r k} & \underline{s}_k \end{pmatrix} \right]^T$$

Expression (3.14) can be written

$$\sum_{k=0}^{N} \sum_{\alpha=0}^{1} \underline{y}_\alpha^k = C_{dd} \sum_{k=0}^{N} \sum_{\alpha=0}^{1} \underline{x}_\alpha^k = \sum_{k=0}^{N} \sum_{\alpha=0}^{1} C_{dd}\, \underline{x}_\alpha^k \tag{3.18}$$

Since the product $C_{dd}\, \underline{x}_\alpha^k$ is another vector $\underline{y}_\alpha^k$ of the same "frequency" k and the same α as $\underline{x}_\alpha^k$, expression (3.18) can be separated into $2N$ or $2N+1$ systems of equations:

$$\underline{y}_\alpha^0 = C_{dd}\, \underline{x}_\alpha^0, \quad \underline{y}_\alpha^1 = C_{dd}\, \underline{x}_\alpha^1 \cdots \underline{y}_\alpha^k = C_{dd}\, \underline{x}_\alpha^k \cdots \underline{y}_\alpha^N = C_{dd}\, \underline{x}_\alpha^N, \quad \alpha = 0, 1 \tag{3.19-a}$$

In turn, solving these systems is the same as finding the solutions to

$$\begin{cases} \underline{v}_\alpha(0) = R(0)\, \underline{\chi}_\alpha(0), & \alpha = 0 \\ \underline{v}_\alpha(1) = R(1)\, \underline{\chi}_\alpha(1) \cdots \underline{v}_\alpha(k) = R(k)\, \underline{\chi}_\alpha(k) \cdots \underline{v}_\alpha(N) = R(N)\, \underline{\chi}_\alpha(N) \end{cases} \tag{3.19-b}$$

where

$$\underline{\chi}_\alpha(k) = \left[\begin{pmatrix} c_{0k} \\ 0 \end{pmatrix} \begin{pmatrix} c_{1k} \\ s_{1k} \end{pmatrix} \cdots \begin{pmatrix} c_{ik} \\ s_{ik} \end{pmatrix} \cdots \begin{pmatrix} c_{N_r k} \\ s_{N_r k} \end{pmatrix} \right]^T \tag{3.20-a}$$

and

$$\underline{v}_\alpha(k) = \left[\begin{pmatrix} m_{0k} \\ 0 \end{pmatrix} \begin{pmatrix} m_{1k} \\ n_{1k} \end{pmatrix} \cdots \begin{pmatrix} m_{ik} \\ n_{ik} \end{pmatrix} \cdots \begin{pmatrix} m_{N_r k} \\ n_{N_r k} \end{pmatrix} \right]^T \tag{3.20-b}$$

3.3 An Algorithm for Solving the Normal Equations $CF^T = C_{sz}^T$

The optimal estimator matrix F relating the minimum variance estimates to the data can be obtained by solving the normal matrix equations (1.5) column by column:

$$C_{dd}\, \underline{f}^{\ell} = \underline{c}_{sz}^{\ell} \tag{3.21}$$

where $\underline{f}^{\ell}$ is the transpose of the ℓth row of F and $\underline{c}_{sz}^{\ell}$ is that of the ℓth row of C_{sz}. So it is necessary to solve N_s systems of equations like (3.21). From the previous section's analysis it is clear that this can be done by decomposing (3.21) into (at most) N_c vectors of dimension N_r such as $\underline{y}_{\alpha}^{k}$ in (3.19-a), and then solving the corresponding systems $\underline{\upsilon}_{\alpha}(k) = R(k)\, \underline{\chi}_{\alpha}(k)$ of (3.19-b). It is much easier to work with the $N_r \times N_r$ matrices $R(k)$ than with the $N_c N_r \times N_c N_r$ matrix C_{dd}, even if there are $N + 1 \simeq N_c/2$ of the smaller matrices. The whole procedure can be described as follows:

Part I

a) Form all $R(k)$ matrices $(0 \le k \le N)$, by Fourier analysis of the first row in every submatrix C_{ij} of C_{dd} (expressions (3.1) and (3.2).

b) Find the corresponding $R(k)^{-1}$, and save all pairs $(R(k), R(k)^{-1})$ on tape or disk if the same type of data, sampled on the same grid, is likely to be used in future estimates.

Part II

c) Decompose the ℓth "right hand side" c_{sz} by Fourier analysis of its N_r partitions, as in expressions (3.15) through (3.17).

d) Form the N_c "equivalent right hand side vectors" $\underline{\upsilon}_{\alpha}(k)$ according to (3.20-b), and solve the corresponding N_c equations (3.19-b) to obtain the "equivalent solution vectors" $\underline{\chi}_{\alpha}(k)$ as in (3.20-a).

e) Use the N_c equivalent solution vectors to generate, by Fourier synthesis (expressions (3.16-a) and (3.17-a)) the ℓth row of F.

f) Repeat steps (c) through (e) for every column in C_{sz}, until all the rows of F have been found.

An alternative to inverting the $R(k)$'s is to generate the pseudoinverse (equivalent to $R(k)^{-1}$ when this exists) of each $R(k)$ by Conjugate Gradients. This method, described in Luenberger (1969), generates a series of N_r-vectors $\underline{v}^i$ that are conjugate directions of $R(k)$: $\underline{v}^{iT} R(k)\, \underline{v}^j = 0$, $i \neq j$, and that can be used to form the pseudoinverse:

$$R(k)^{\dagger} = \sum_{i=0}^{R} (\underline{v}^{iT} R(k)\, \underline{v}^i)^{-1}\, \underline{v}^i\, \underline{v}^{iT}, \quad R = \text{rank}\,[R(k)] \tag{3.22}$$

In cases when C_{dd}^{-1} (and, therefore, at least one $R(k)^{-1}$) does not exist, the normal equations have, nonetheless, solution. This is because, being covariance matrices, C_{sz}^{T} is always in the span of C_{dd}. This means that there is always an exact solution to $\underline{v}_\alpha(k) = R(k)\,\underline{x}_\alpha(k)$ that can be obtained

$$\underline{v}_\alpha(k) = R^{\dagger}(k)\,\underline{x}_\alpha(k) \tag{3.23}$$

This idea has been used successfully to get the results in paragraph 5.5.

3.4 Equatorial Symmetry

If every row in the grid has a counterpart in the opposite hemisphere and both are at the same spherical distance from their poles, then the grid has "equatorial symmetry". The equator itself (an equatorial band, in the case of blocks) can be one of the rows.

If the grid has equatorial symmetry, then C_{dd} is persymmetric: two submatrices C_{ij} are equal[1] if they are symmetrically situated with respect to the main diagonal (ordinary symmetry) or the main antidiagonal. More generally, one can permute the i th row with the i th column, or the i th row with the $N_d + 1 - i$ th column, in a persymmetric matrix of dimension N_d, without modifying the matrix. Since $R(k)$ is formed by taking one Fourier coefficient a_k^{ij} from each C_{ij}, it follows that $R(k)$ is also persymmetric.

An important property of persymmetric matrices, from the point of view of this study, is the following:

(1) If $\underline{v}$ is an "even" vector:

$$\underline{v} = [\,v_1\, v_2 \ldots\, v_i \ldots\, v_{N_r}\,]^T \tag{3.24}$$

$$v_1 = v_{N_r};\ v_2 = v_{N_r-1} \ldots\ v_i = v_{N_r-i} \quad \text{for } 0 < i \le N_H = \begin{cases} N_r/2 & \text{if } N_r \text{ is even} \\ (N_r-1)/2 & \text{if } N_r \text{ is odd} \end{cases}$$

(2) Or if $\underline{v}$ is an "odd" vector:

$$\underline{v} = [\,v_1\, v_2 \ldots\, v_i \ldots\, v_{N_r}\,]^T$$

$$v_1 = -v_{N_r};\ v_2 = -v_{N_r-1} \ldots v_i = -v_{N_r-i} \quad \text{for } 0 \le i \le N_H \tag{3.25}$$

then the product of $\underline{v}$ by a persymmetric matrix is also "even" or "odd", respectively. In other words: multiplication by a persymmetric matrix preserves the "parity" of the vector. Any vector $\underline{b}$ can be decomposed into an "even" part $\underline{b}_{\beta=0}$ and an "odd" part $\underline{b}_{\beta=1}$:

[1] Remember that all C_{ij} are symmetrical and, being Toeplitz circulant, are also persymmetrical.

$$\underline{b}_{\beta=0} = [b_1^0\ b_2^0 \ldots b_i^0 \ldots b_{N_r}^0]^T \qquad b_i^0 = \tfrac{1}{2}(b_i + b_{-i+N_r+1}),\ (b_{N_H+1}^0 = b_{N_H+1} \text{ if } N_r \text{ odd})$$

$$\underline{b}_{\beta=1} = [b_1^1\ b_2^1 \ldots b_i^1 \ldots b_{N_r}^1]^T \qquad b_i^1 = \tfrac{1}{2}(b_i - b_{-i+N_r+1}) \qquad (3.26)$$

Therefore, an equation of the type

$$\underline{v}_\alpha(k) = R(k)\,\underline{x}_\alpha(k)$$

can be separated into two independent equations:

$$\underline{v}^{(k)}_{\alpha\beta=0} = R(k)\,\underline{x}^{(k)}_{\alpha\beta=0} \qquad (3.27\text{-a})$$

$$\underline{v}^{(k)}_{\alpha\beta=1} = R(k)\,\underline{x}^{(k)}_{\alpha\beta=1} \qquad (3.27\text{-b})$$

where $\beta=0$ indicates "even", and $\beta=1$ "odd". The solution (3.27-a) must be "even", while that to (3.27-b) must be "odd", so the actual number of "degrees of freedom" is $N_H \simeq$ number of unknowns/2. When half of the unknowns are found, the other half must have the same or opposite values. Therefore, only the first N_H equations in (3.27-a,b) are needed to solve the system. Instead of the original system, one can solve the equivalent:

$$\tilde{\underline{v}}^{(k)}_{\alpha\beta} = \tilde{R}_\beta(k)\,\tilde{\underline{x}}^{(k)}_{\alpha\beta} \qquad (3.28\text{-a})$$

where $\tilde{\underline{v}}^{(k)}_{\alpha\beta}$ and $\tilde{\underline{x}}^{(k)}_{\alpha\beta}$ have dimension N_H, $\tilde{R}_\beta(k)$ is $N_H \times N_H$ and has elements

$$\tilde{r}^{k\beta}_{mn} = r^k_{mn} + r^k_{mN_r-n+1}(-1)^\beta \qquad (3.28\text{-b})$$

with $1 \le n \le N_H$ and $1 \le m \le N_H$, where $\tilde{\underline{v}}^{(k)}_{\alpha\beta}$ contains the first N_H elements in $\underline{v}^{(k)}_{\alpha\beta}$ and $\tilde{\underline{x}}^{(k)}_{\alpha\beta}$ the first N_H in $\underline{x}^{(k)}_{\alpha\beta}$. There are twice as many equations such as (3.28) than there are equations like (3.19), but the reduction in size of the matrices by half brings a considerable increase in efficiency.

4. Computing

This section considers the implementation of the algorithm for grided data from the point of view of efficiency and of reliability of results. Also certain numerical stability matters are treated.

4.1 Setting up the Matrix $C_{dd} = C_{zz} + D$

Usually D is a diagonal matrix, so calculating its contribution to C_{dd} is trivial. If this is not the case ("colored noise"), this matrix is handled in the same way as C_{zz}. C_{zz} contains all the covariances of the signal in the data: because the symmetries in the grid are reflected in the structure of C_{zz}, it is not necesaary to compute every one of them. C_{zz} is symmetrical, so only half

of its elements have to be found. Every C_{ij} block is Toeplitz circulant, so only the first row has to be known, and this first row is "even", as already explained, so only half of its elements are different. If there is equatorial symmetry, then C_{zz} is persymmetric, and this reduces the number of distinct elements by half once more. The number of covariances in C_{zz} is $(N_c N_r)^2$: after all symmetries are considered, only $N_c N_r^2/4$ have to be computed if the grid is not symmetric, and only $N_c N_r^2/8$ if it is. Computing each covariance should take the same amount of operations, so the central processor time needed is porportional to the number of distinct elements. This means that a matrix such as C_{zz} requires $4N_c$ times less to be set up than an ordinary symmetric matrix of the same dimension, and $8N_c$ times less if the grid is symmetrical. In the case of a regular $1° \times 1°$ grid with 64800 points (blocks) there is a reduction in effort of the order of 1400 times. Furthermore, the existence of redundancy in the elements of C_{dd} can be used to decrease the storage requirements for this matrix, that would otherwise be truly enormous even with moderately large data sets.

4.2 Solving the Normal Equations

Solving a system of equations requires a number of operations proportional to the cube of the number of unknowns. Assuming that the same method were used to solve the original equation for one row of F

$$\underline{f}^{\ell} = C_{dd}^{-1}\, \underline{c}_{sz}^{\ell}$$

that is used to solve each of the equations

$$\underline{v}_{\alpha}(k) = R(k)\, \underline{x}_{\alpha}(k) \quad \text{or} \quad \tilde{\underline{v}}_{\alpha\beta}^{(k)} = \tilde{R}_{\beta}(k)\, \tilde{\underline{x}}_{\alpha\beta}^{(k)}$$

as the case may be, then the total saving in computing time due to the structure of C_{dd} is of the order of N_c^2. The use of a symmetric grid increases efficiency by a factor of four. The additional time needed to do the "Fourier analysis" of the right hand sides or the "Fourier synthesis" of the solution vectors is quite negligible, even for large numbers of data points, thanks to the existance of very efficient algorithms, particularly in the case when N_c is a power of 2.[1] In the case of the regular $1° \times 1°$ grid, with $N_c = 360$, the reduction of computing time is of the order of 130000.

Computing the covariances that form the C_{dd} matrix, even after all symmetries have been fully exploited, remains no trivial task if the data points (blocks) are very numerous. In the case of point data this situation is helped by the existence of closed covariance expressions such as those found by Tscherning and Rapp (1974) for isotropic functions. In the case of block data, as paragraph 5.3 shows, the covariances of area means are "area means of area means of covariances," involving double area integrals over the various blocks. Numerical integration could be used as an approximation, but this would require a major

[1] Fast Fourier Transforms.

operation if the dimension of C_{dd} is large. It would be most desirable to have closed expressions (or convenient approximations) for area means' covariances, but they are not known to this author.[1] Reductions in computer time, while valuable in themselves, are not the only important gain: fewer computations mean less rounding errors accumulation, and more reliable results.

All the properties of the covariance matrix mentioned so far apply, in the case of isotropic and other covariances, not only on the sphere but also on any surface of revolution, as long as the "rows" are defined by circles perpendicular to the axis of rotation. Such surfaces include: the cylinder, the cone, and the geodesist's old friend, the oblate spheroid. The same structure arises from concentric rings on a plane and, of course, from the regular sampling of a circumference. Equispaced sampling along a straight line results in a Toeplitz matrix, different from a Toeplitz circulant one in that the last element of a row is usually "lost" because a different number appears as the first element in the following row, while all other elements are shifted one place to the right as before. Equispaced sampling on a rectangular grid in the plane produces Toeplitz block matrices of Toeplitz blocks. All Toeplitz matrices can be set up and inverted efficiently, with approximately (dimension)^2 operations per inversion. This is also the case with the type of block matrices discussed in this paper, as already shown. The properties of Toeplitz-type (and closely related Hankel-type) matrices have been used to devise algorithms for minimum variance prediction and filtering on the real line (time domain) and on the plane. Examples of the first application are the algorithms of Levinson (1947) and Trench (1964). For the plane there is an interesting method due to Justice (1977). In Geodesy there has been a recent application of Toeplitz matrices to the prediction of ocean gravity anomalies from satellite altimetry, by Eren (1979). Besides what might be called "outright" Toeplitz matrices (circulant, block, or plain) there are "near Toeplitz" matrices and operators which have, to a lesser degree, some of the advantages considered here. Such structures have been studied by Kailath (1975) among others: it was at a talk delivered by him at the University of New South Wales, in early 1976, that the author of this paper first became aware of the many uses of Toeplitz matrices.

4.3 Numerical Stability

The poles (circular blocks about the poles) have been excluded by restriction C-2 (C-2') because, in order to partition C_{dd} into N_c^2 $N_r \times N_r$ submatrices C_{ij}, the corresponding measurement would have to be artificially treated as N_c measurements at the same point (block), introducing N_c rows and columns in C_{ii} that are identical, thus making C_{dd} singular. Even with this restriction, rows very close to the pole may create stability problems, as the matrix will tend to become

[1] One approximation, based on the Pellinen "smoothing factors" (rendered into a recursive form by Meissl (1971)), might be acceptable for very fine subdivisions of the sphere, though it requires the use of Legendre harmonic expansions truncated to a high degree.

singular as the rows approach the pole.[1] In particular, small perturbations in the elements of C_{dd} may have serious consequences for the solution of the normals. The author found, when computing the results presented in Example III, where the grid was confined to small circular caps, that interpolating linearly from a table instead of computing each covariance exactly from a closed expression (in order to save time) resulted in a matrix with some negative eigenvalues when the table entries were spaced at more than 0.25km intervals! He finally computed all covariances from the closed expression, and the problem disappeared. For a discussion of the interpolation problem, see Sünkel (1978).

Another way of computing covariances approximately is by truncating their spherical harmonic expansions, which, in the isotropic case,[2] are of the type

$$c_{zz}(\psi) = \sum_{n=2}^{\infty} c_{zz,n} P_n(\cos\psi), \quad c_{sz}(\psi) = \sum_{n=2}^{\infty} c_{sz,n} P_n(\cos\psi) \tag{4.1}$$

at a "sufficiently high" degree N_{max} (usually $N_{max} = 1000$). If the spacing between meridians is $\Delta\lambda < \pi/N_{max}$ then $C_{dd}\, \underline{f}_\alpha^k = 0$, where

$$\underline{f}_\alpha^k = \left[\varphi_1^k\left(\frac{\underline{c}_k}{\underline{s}_k}\right)\varphi_2^k\left(\frac{\underline{c}_k}{\underline{s}_k}\right) \ldots \varphi_1^k\left(\frac{\underline{c}_k}{\underline{s}_k}\right) \ldots \varphi_{N_r}^k\left(\frac{\underline{c}_k}{\underline{s}_k}\right) \right]^T,$$

if $N_{max} < k <$ Integer $(\pi/\Delta\lambda)$. Furthermore, $R(k) = 0$ if $N_{max} < k <$ Integer $(\pi/\Delta\lambda)$. This presents no problem if $c_{sz}(\psi)$ is computed using an expansion truncated to N_{max}, because c_{sz} contains no higher frequencies in its columns, which are in the span of C_{dd}, and the desired solution can be obtained using the non-zero $R(k)$, $0 \le k \le N_{max}$.

4.4 Eigenvector and Eigenvalue Decomposition of C_{dd}

One possibility, when dealing with an ill-conditioned, real symmetric matrix, C, is to decompose it into eigenvectors and eigenvalues:

$$C = \sum_{i=1}^{\text{Rank } C} \lambda_i \underline{\mu}_i \underline{\mu}_i^T$$

(where the $\underline{\mu}_i$ are the eigenvectors and the λ_i eigenvalues) and then form a pseudoinverse

$$C^\dagger = \sum_{q=1}^{k} \lambda_q^{-1} \underline{\mu}_q \underline{\mu}_q^T$$

where the λ_q are those eigenvalues of C that satisfy the condition $\lambda_q > \epsilon > 0$

---

[1] In the case of area means this problem can be alleviated by using equal area grids with blocks of constant longitude span, and latitude spans that increase towards the poles.

[2] P_n is the nth unnormalized Legendre polynomial; $c_{zz,n}$ and $c_{sz,n}$ are the nth degree variance of z and the nth degree covariance between z and s, respectively.

for ϵ "sufficiently small". In other words: the true inverse is "truncated" to that part of its expansion that can be regarded as "sufficiently positive definite". For this and other reasons it is interesting to know the eigenvector/eigenvalue decomposition of C_{dd}.

From Section 3 it follows that if

$$\underline{\mu}(m)_i = [\mu(m)_{i,1} \ldots \mu(m)_{i,N_r}]^T$$

is an eigenvector of $R(m)$ and $\lambda(m)_i$ is the corresponding eigenvalue, then $\lambda(m)_i$ is also an eigenvalue of C_{dd}, and

$$\underline{v}^{\alpha}(m)_i = [\mu(m)_{i,1}\left(\frac{c_\alpha}{s_\alpha}\right) \ldots \mu(m)_{i,N_r}\left(\frac{c_\alpha}{s_\alpha}\right)]^T$$

the associated pair of eigenvectors of C_{dd}. Therefore, the spectral decomposition of the $N_c/2$ matrices $R(m)$ is equivalent to that of C_{dd}. However, it is far easier to decompose $N_c/2$ $N_r \times N_r$ matrices than to do the same to one $N_r N_c \times N_r N_c$ matrix.

4.5 Regularization of the Normal Equations

Sometimes, when a matrix C is too ill-conditioned for the solution to the corresponding system $\underline{y} = C\underline{x}$ to be computed reliably, a simple form of regularization, that gives more stable results to a slightly different problem, consists in solving the system $\underline{y} = (C + \alpha I)\underline{x}$ (where α is a very small, positive constant) instead of the original equations.

The "trick" consists in finding a value of α that stabilizes the results without causing them to depart too much from those for the original system. This can be done with relative ease if a spectral decomposition for C is available. In the case at hand, the normal equations will have, after this regularization, a "normal matrix" $C_{dd} + \alpha I_{(N_d \times N_d)}$. This can be "Fourier analyzed" as before, yielding the $N_c/2$ matrices

$$R'(m) = R(m) + \alpha I_{(N_r \times N_r)} \qquad (4.3)$$

so the regularization of C_{dd} implies that of each $R(m)$, which can be handled as in Tikhonov and Arsenin (1977).

4.6 Grids of Higher Symmetry

The high efficiency in setting up and solving the equation $C_{dd}\, F^T = C_s^T{}_z$ made possible by the structure of C_{dd} raises the question of the possible existence of partitions of the sphere that generate even stronger structures. The answer is yes, and, as examples, consider: a single "row", two "rows" sym-

metrical with the equator, and the vertices of the five regular (and of the 13 semi-regular) solids. In all these cases the matrix C_{dd} is a block Toeplitz circulant matrix of circulant blocks, while the matrices considered so far were simply block matrices of circulant blocks.[1] With the exception of the second arrangement, the size of the C_{ij} is 1 x 1, which reduces the whole matrix to an ordinary Toeplitz circulant matrix, the setting up and inverting of which is almost trivial. Are there such grids with large numbers of nodes (blocks) evenly distributed over the whole sphere? The answer to this question is not known to the author, but its importance can be appreciated by the reader. Paulik (1976) has published a theorem containing a sufficient condition for the existence of this type of grid, as well as a constructive principle, linking its existence to that of pairs of commuting, nontrivial, 3 x 3 orthogonal matrices. Whether some of these pairs correspond to dense grids is another matter.

If C_{dd} is block Toeplitz circulant of circulant blocks, then the elements in each row (column) are the same, only their order changes. In the case of isotropic covariances, this means that the set of distances from any data point (block) to all the others must be independent of the data point (block) chosen. Clearly this is a necessary condition.

5. Examples

This section illustrates the application of the method to spherical harmonic analysis of gridded point data and of area averages, and to estimating disturbing potential from gravity anomalies.

5.1 Spherical Harmonic Analysis of Point Data

Spherical harmonic analysis is to data distributed on a sphere, what Fourier analysis is to data on the line or on the plane. Not only does it provide greater insight into the properties of the information available, its statistics, and its relationships to other signals (see Kaula, 1967), but it also allows the highly efficient computation of convolutions. Such is the case of a function[2]

$$h(\varphi,\lambda) = \sum_{n=0}^{\infty} \sum_{m=0}^{n} \bar{P}_{nm}(\varphi)(\bar{c}_{nm}\cos m\lambda + \bar{s}_{nm}\sin m\lambda) \qquad (5.1)$$

that is transformed according to

---

[1] If the data were partioned by meridians instead of by parallels, C_{dd} would be a block Toeplitz circulant matrix, but the blocks themselves would not be Toeplitz.

[2] $\bar{P}_{nm}$ is the associated Legendre function of the first kind, of order n and degree m (normalized); $\bar{c}_{nm}$ and $\bar{s}_{nm}$ are fully normalized coefficients; P_n is the unnormalized Legendre polynomial of degree n ; $\int_\sigma d\sigma$ indicates an area integral over the whole sphere; ψ is the spherical distance from (φ,λ) to (φ',λ') (see, for instance, Hobson (1965)). $\bar{P}_{nm}(\varphi)$ is shorthand for $\bar{P}_{nm}(\sin\varphi)$.

$$u(\varphi,\lambda) = \frac{1}{4\pi}\int_\sigma S(\psi)\, h(\varphi',\lambda')\, d\sigma \quad \text{(as in Stokes' formula)} \tag{5.2}$$

for some $S(\psi) = \sum_{n=0}^{\infty} k_n (2n+1) P_n(\psi)$

in which case

$$u(\varphi,\lambda) = \sum_{n=0}^{\infty}\sum_{m=0}^{n} k_n \bar{P}_{nm}(\varphi)\,[\bar{c}_{nm}\cos m\lambda + \bar{s}_{nm}\sin m\lambda] \tag{5.3}$$

Computing (5.2) by numerical quadratures is far more laborious, if u is required at many points, than using (5.3) truncated to a high degree and order, if the coefficients $\bar{c}_{nm}$, $\bar{s}_{nm}$ are known. Finding these coefficients accurately and with a minimum of computations is a very desirable goal: a number of studies have been published in recent years on the "correct" way of analyzing data, particularly when given in the form of area means (see, for instance, Rapp (1977) and Katsambalos (1979)). Much of the effort has been concentrated on computing the coefficients from the expressions

$$\bar{c}_{nm} = \frac{1}{4\pi}\int_\sigma \bar{R}_{nm}(\varphi,\lambda)\, h(\varphi,\lambda)\, d\sigma \tag{5.4-a}$$

$$\bar{s}_{nm} = \frac{1}{4\pi}\int_\sigma \bar{S}_{nm}(\varphi,\lambda)\, h(\varphi,\lambda)\, d\sigma \tag{5.4-b}$$

based on the orthogonality of the harmonics $\bar{R}_{nm}(\varphi,\lambda) = \bar{P}_{nm}(\sin\varphi)\cos m\lambda$ and $S_{nm}(\varphi,\lambda) = P_{nm}(\sin\varphi)\sin m\lambda$ on the sphere, using numerical quadratures. Such approaches can be very efficiently implemented: in 1976 C. Rizos and the author wrote Fortran programs for harmonic analysis and synthesis. As an example, one of those programs took 1.3 minutes to generate a set of $\bar{c}_{nm}$'s and $\bar{s}_{nm}$'s complete to degree and order 180 from 64800 $1^\circ \times 1^\circ$ area means (gravity anomalies) in the AMDHAL 470V/6-II computer of the Ohio State University.

Because the data is sampled, there is usually not enough of it to estimate the coefficients exactly: the resulting error is known as aliasing, and it depends both on the data distribution and on the numerical technique used. Moreover, the data usually contains spurious signals, measurement errors for instance, that also affect the results. A way of computing harmonic coefficients, minimizing the effect of noise and aliasing simultaneously, has been described by Rummel (1976) and by Sjöberg (1978). The idea is to estimate the coefficients using the minimum variance method, or collocation, which involves solving the normal equations $C_{dd}\, F^T = C_{sz}^T$ for this particular problem. Under the restrictions listed earlier on, which means, for instance, using an ordinary regular grid and isotropic covariances as defined by (4.1) with all data on the same sphere, the C_{dd} matrix has the advantageous properties mentioned so far, and can be treated accordingly. What of the C_{sz} matrix? The $\bar{c}_{nm}$ and $\bar{s}_{nm}$ are functions of the system of coordinates chosen: rotating the λ origin, and shifting the poles change their values. If φ and λ are the coordinates of the shifted N pole with respect to some fixed system, then $\bar{c}_{nm} \equiv \bar{c}_{nm}(\varphi,\lambda)$ and $\bar{s}_{nm} \equiv \bar{s}_{nm}(\varphi,\lambda)$ can be regarded as functions of the pole coordinates: ordinary functions of φ

[1] At the Department of Geodesy, School of Surveying, The University of New South Wales, Australia.

and λ to be estimated at the "North pole" of the grid. Assuming that $M\{\ \}$ is the isotropic operator, using (5.4) and the orthogonality properties of surface harmonics:

$$\begin{aligned} M\{\ \bar{c}_{nm}\, h(\varphi_i,\lambda_j)\} &= \frac{1}{4\pi} M\Big\{ \int_\sigma \bar{R}_{nm}(\varphi,\lambda)\, h(\varphi,\lambda)\ h(\varphi_i,\lambda_j)\, d\sigma \Big\} \\ &= \frac{1}{4\pi} \int_\sigma \bar{R}_{nm}(\varphi,\lambda)\ M\{\ h(\varphi,\lambda)\ h(\varphi_i,\lambda_j)\ \}\, d\sigma \\ &= \frac{1}{4\pi} \int_\sigma \bar{R}_{nm}(\varphi,\lambda) \sum_{n=0}^{\infty} c_n\, P_0(\psi_{PQ})\, d\sigma, \quad (P=(\varphi,\lambda),\ Q=(\varphi_i,\lambda_j)) \\ &= \frac{c_n}{(2n+1)} \bar{R}_{nm}(\varphi_i,\lambda_j) \end{aligned}$$

Similarly,

$$M\{\ \bar{s}_{nm}\, h(\varphi_i,\lambda_j)\ \} = \frac{s_n}{(2n+1)} \bar{S}_{nm}(\varphi_i,\lambda_j)\,,$$

So

$$M\left\{ \begin{pmatrix} \bar{c}_{nm} \\ \bar{s}_{nm} \end{pmatrix}\ h(\varphi_i,\lambda_j) \right\} = \frac{c_n}{(2n+1)} \bar{P}_{nm}(\sin\varphi_i) \begin{pmatrix} \cos \\ \sin \end{pmatrix} \frac{2\pi}{N}\ mj \qquad (5.5)$$

From (5.5), it is clear that the columns of C_{sz}^T are already separated in frequency, with m taking here the place of k for convenience in notation. In the case of equatorial symmetry $\underline{v}^n_{\alpha=0}(m) = [\bar{P}_{nm}(\varphi_1), \bar{P}_{nm}(\varphi_2), \ldots, \bar{P}_{nm}(\varphi_{N_r})]^T$ and $\underline{v}^n_{\alpha=1}(m) = [\bar{P}_{nm}(\varphi_1), \ldots, \bar{P}_{nm}(\varphi_{N_r})]^T$ are identical and either "even" or "odd" vectors, because the P_{nm} are even functions of φ if n-m is even, and odd functions if n-m is odd. In any event, $\underline{v}_{\alpha=0}(m) = \underline{v}_{\alpha=1}(m)$ regardless of symmetry. The complete decomposition of the columns of C_{sz}^T (right hand sides) is immediate: all that has to be computed are the values of $P_{nm}(\varphi_i)$ for $1 \le i \le N_r$, for which there are simple recursive formulas by order and by degree. After solving the reduced equations:

$$\underline{v}^n(m) = R(m)\, \underline{x}^n(m) \quad \text{or} \quad \underline{\tilde{v}}^n_\beta(m) = \tilde{R}(m)\, \underline{\tilde{x}}^n_\beta(m)$$

(depending on whether the grid is symmetric or not w.r.t. the equator), where the subscript α is superfluous and has been dropped, the "synthesis" of the corresponding row in F is also immediate, say

$$\underline{f}^{nm}_{\alpha\beta} = [\tilde{x}^n_{\beta,1,n}(\underline{c}_m), \tilde{x}^n_{\beta,2,n}(\underline{c}_m), \ldots, \tilde{x}^n_{\beta,N_r,n}(\underline{c}_m)]^T$$

in the equatorially symmetric case, where $\underline{f}^{nm\,T}_{\alpha\beta}$ is the row in F corresponding to $\underline{\hat{c}}_{nm}$ or $\underline{\hat{s}}_{nm}$. Once the $\underline{\tilde{x}}^n_\beta(m)$ are known, the estimates are obtained as follows (again in the symmetric case).

$$\begin{aligned} \begin{pmatrix} \hat{c}^{nm} \\ \hat{s}_{nm} \end{pmatrix} = \underline{f}^{nm\,T}_{\alpha\beta}\, \underline{d} &= \sum_{i=1}^{N_M} \tilde{x}^n_{\beta,i,n} \begin{pmatrix} c_n^T \\ s_n^T \end{pmatrix} \underline{d}_i + \sum_{i=N_M+1}^{N_r} \tilde{x}^n_{\beta,i,n} \begin{pmatrix} c_n^T \\ s_n^T \end{pmatrix} \underline{d}_i\, (-1)^\beta \\ &= \sum_{i=1}^{N_M} \tilde{x}^n_{\beta,i,n} \begin{pmatrix} a_i^m \\ b_i^m \end{pmatrix} + \sum_{i=N_M+1}^{N_r} \tilde{x}^n_{\beta,i,n} \begin{pmatrix} a_i^m \\ b_i^m \end{pmatrix} (-1)^\beta, \quad t = N_r+1-i \qquad (5.6) \end{aligned}$$

$$a_i^m = \underline{c}_m^T \underline{d}_i = \sum_{j=0}^{N_c-1} \cos\frac{2\pi mj}{N_c}\, d_{ij} \qquad (5.7\text{-a})$$

$$b_i^m = \underline{s}_m^T \underline{d}_i = \sum_{j=0}^{N_c-1} \sin\frac{2\pi mj}{N_c}\, d_{ij} \qquad (5.7\text{-b})$$

Expressions (5.7) represent the Fourier analysis of each N_c partition of the $N_r N_c$-data vector $\underline{d}$. So the algorithm for getting $\hat{\bar{c}}_{nm}$ and $\hat{\bar{s}}_{nm}$, once the values of the $\underline{X}\beta^n(m)$ are known, is as follows:

a) Find the $\begin{pmatrix} a_i^m \\ b_i^m \end{pmatrix}$ by Fourier analysis of the $\underline{d}_i$;

b) Use the $\underline{X}\beta^n(m)$ and $\begin{pmatrix} a_i^m \\ b_i^m \end{pmatrix}$ as in expressions (5.6) and (5.7) to find the estimated $\bar{c}_{nm}$ and $\bar{s}_{nm}$.

For reasons explained in the next paragraph, m is to be limited to the range $0 \le m \le N = \begin{cases} N_c/2 & \text{if } N_c \text{ is even} \\ (N_c-1)/2 & \text{if } N_c \text{ is odd.} \end{cases}$

Steps (a) and (b) can be performed very efficiently, even for large numbers of data points, because of the power of the Fourier algorithms available at present. Clearly, all $\hat{\bar{c}}_{nm}$ and $\hat{\bar{s}}_{nm}$ of the same <u>order</u> m can be found quite independently from the rest of the coefficients. This <u>separability in order</u> is also present in the calculation of coefficients by <u>least squares adjustment</u> on a regular grid. Such adjustment is based on the "observation equations"

$$h(\varphi_i, \lambda_j) = \sum_{n=0}^{N} \sum_{m=0}^{n} \begin{pmatrix} \bar{R}_{nm} \\ \bar{S}_{nm} \end{pmatrix} (\varphi_i, \lambda_j) \begin{pmatrix} \bar{c}_{nm} \\ \bar{s}_{nm} \end{pmatrix}$$

In this technique, as in collocation, there is also separation by frequency and parity. When $n = N$ and $m = N$ are the largest degree and order present in the data, the least squares method yields perfect estimates (assuming no noise). Collocation, being a minimum variance technique, also gives perfect estimates when applied to such data. An important difference between the two procedures is that, for

$$h(\varphi, \lambda) = \sum_{n=0}^{\infty} \sum_{m=0}^{n} [\bar{c}_{nm} \bar{R}_{nm}(\varphi, \lambda) + \bar{s}_{nm} \bar{S}_{nm}(\varphi, \lambda)]$$

least squares has no control on the aliasing, while collocation minimizes it (provided the appropriate covariances are used).

5.2 Aliasing

When data is noiseless, the error in $\begin{pmatrix} \hat{\bar{c}}_{nm} \\ \hat{\bar{s}}_{nm} \end{pmatrix}$ is a function of higher degree and order coefficients. This type of error is commonly known as aliasing: high frequency waves become indiscernible from lower frequency ones because of the sampling. To understand this problem, aliasing can be considered first <u>by degree</u> and then <u>by order</u>.

I - <u>Higher degrees</u>: There are N_r samples in latitude, so there are at most N_r independent columns in C_{sz}^r of the form

$$\underline{c}_{sz,\alpha}^{nm} = [\bar{P}_{nm}(\varphi_1)\begin{pmatrix} \underline{c}_m \\ \underline{s}_m \end{pmatrix}, \bar{P}_{nm}(\varphi_2)\begin{pmatrix} \underline{c}_m \\ \underline{s}_m \end{pmatrix}, \ldots, \bar{P}_{nm}(\varphi_{N_r})\begin{pmatrix} \underline{c}_m \\ \underline{s}_m \end{pmatrix}]^r$$

for any given m. Therefore, if $n > N_r$,

$$\underline{c}_{sz,\alpha}^{nm} = \sum_{i=1}^{N_r} a_i \underline{c}_{sz,\alpha}^{im}$$

for some real numbers a_i. If C_{dd} is invertible, F has the same rank as C_{sz}, and there are only N_r independent $\underline{f}_\alpha = C_{dd}^{-1} \underline{c}_{sz,\alpha}^{nm}$. If $n > N_r$

$$\underline{f}_\alpha^{nm} = C_{dd}^{-1} \sum_{i=1}^{N_r} \underline{c}_{sz,\alpha}^{im} a_i = \sum_{i=1}^{N_r} C_{dd}^{-1} \underline{c}_{sz,\alpha}^{im} a_i = \sum_{i=1}^{N_r} \underline{f}_\alpha^{im} a_i$$

and

$$\begin{pmatrix} \hat{C}_{nm} \\ \hat{S}_{nm} \end{pmatrix} = \underline{f}_\alpha^{nm\,T} \underline{d} = \sum_{i=1}^{N_r} a_i \underline{f}_\alpha^{im\,T} \underline{d} = \sum_{i=1}^{N_r} a_i \begin{pmatrix} \hat{C}_{im} \\ \hat{S}_{im} \end{pmatrix}$$

or, after several more steps,

$$\begin{pmatrix} \hat{C}_{im} \\ \hat{S}_{im} \end{pmatrix} = \sum_{n=N_r+1}^{\infty} \alpha_n \begin{pmatrix} \bar{C}_{nm} \\ \bar{S}_{nm} \end{pmatrix} \qquad (5.8)$$

for some real numbers α_n. Expression (5.8) indicates a dependence of the estimated coefficients $\begin{pmatrix} \hat{C}_{im} \\ \hat{S}_{im} \end{pmatrix}$ on coefficients of higher order. The actual coefficients are quite independent from each other.

II - Higher orders: There are N_c samples per row, so only $N_c/2$ [1] terms in $\cos m\lambda_j$ and $N_c/2$ in $\sin m\lambda_j$ (defined in (5.5)) can be independent simultaneously. This means that terms of different frequency are lumped together, as shown by the trigonometric relationship:

$$\begin{pmatrix} \cos \\ \sin \end{pmatrix} \frac{2\pi}{N_c} mq = \begin{pmatrix} \cos \\ \sin \end{pmatrix} \frac{2\pi}{N_c} (m + N_c k) q$$

where $k = 0, 1, 2, \ldots$. Consequently, both the rows in C_{sz} and in F corresponding to $\begin{pmatrix} \bar{C}_{nm} \\ \bar{S}_{nm} \end{pmatrix}$ with $m > N_c/2$ will have the form

$$[\eta_1 \begin{pmatrix} \underline{c}_m \\ \underline{s}_m \end{pmatrix} \eta_2 \begin{pmatrix} \underline{c}_m \\ \underline{s}_m \end{pmatrix} \cdots \eta_i \begin{pmatrix} \underline{c}_m \\ \underline{s}_m \end{pmatrix} \cdots \eta_{N_r} \begin{pmatrix} \underline{c}_m \\ \underline{s}_m \end{pmatrix}]^T = [\eta_1 \begin{pmatrix} \underline{c}_{m'} \\ \underline{s}_{m'} \end{pmatrix} \eta_2 \begin{pmatrix} \underline{c}_{m'} \\ \underline{s}_{m'} \end{pmatrix} \cdots \eta_n \begin{pmatrix} \underline{c}_{m'} \\ \underline{s}_{m'} \end{pmatrix} \cdots \eta_{N_r} \begin{pmatrix} \underline{c}_{m'} \\ \underline{s}_{m'} \end{pmatrix}]^T$$

(where $m = m' + (N_c k)/2$) and will be linear combinations of the N_r independent columns with the same lower order m'. It follows that there is dependency between coefficients of order m' and those of order

$$m = m' + N_c,\ m' + 2N_c,\ m' + 3N_c, \ldots,\ m' + kN_c, \ldots$$

From (I) and (II) it is clear that estimates of degree $n < N_r$ and order $m < N_c/2$ are functions of higher frequency terms. This dependency usually gets worse as the upper limits are approached by n and m. In regular partitions, where $\Delta\varphi = \Delta\lambda$, so $N_r \simeq N_c/2$, this common upper limit for degree and order is usually referred to as the "Nyquist frequency" $N_c/2$, after its time series analogue.

Aliasing is a problem present in all forms of harmonic analysis, including that on the sphere. Some methods, however, are less affected by it than others. For noiseless data, aliasing is the error in the estimates of the coefficients. So, if the mean square error in each coefficient is the criterion for measuring aliasing

[1] $N_c/2$ is only correct within one unit of the actual number. "$N_c/2$" is used here and in the remainder as a simplification.

for that coefficient, then for a given class of functions $h(\varphi,\lambda)$ with covariance

$$M\{h(P)\,h(Q)\} = \sum_{n=0}^{\infty} c_n P_n(\psi_{PQ})$$

or degree variances (power spectrum) $\sigma_n = c_n$, minimum variance analysis (or collocation) is the procedure with the least possible aliasing.

5.3 Spherical Harmonic Analysis from Area Means

Sjöberg (1978) has derived expressions for the isotropic covariances of area means of gravity anomalies over blocks of the type shown in Figure 2.1-b, and for the covariances between such area means and the normalized harmonic coefficients of the signal before averaging. Generalizing those formulas to area means $\bar{h}$ of a function h, we have

$$M\{\bar{h}_i \bar{h}_j\} = \frac{1}{\Delta\sigma_i \Delta\sigma_j} \int_{\Delta\sigma_i} \int_{\Delta\sigma_j} c_{hh}(\psi)\, d\sigma\, d\sigma \qquad (5.9)$$

where $\Delta\sigma_i, \Delta\sigma_j$ are the areas of the ith and jth blocks, and

$$M\left\{\begin{pmatrix}\bar{c}_{nm}\\ \bar{s}_{nm}\end{pmatrix}, \bar{h}_j\right\} = \frac{c_n}{(2n+1)} \frac{\int_{\varphi_{S_j}}^{\varphi_{N_j}} \bar{P}_{nm}(\psi)\cos\varphi\, d\varphi}{(\sin\varphi_{N_j} - \sin\varphi_{S_j})} \times$$

$$\times \quad \begin{cases} 1 & \text{if } m = 0,\ \alpha = 0 \\ (1/m)\left[a_m(\Delta\lambda)\begin{pmatrix}\cos\\ \sin\end{pmatrix} m\lambda_{W_j} + (-1)^{\alpha} b_m(\Delta\lambda)\begin{pmatrix}\sin\\ \cos\end{pmatrix} m\lambda_{W_j}\right] \end{cases}$$

where $a_m(\Delta\lambda) = \sin m\Delta\lambda$
$b_m(\Delta\lambda) = (1 - \cos m\Delta\lambda)$
$\lambda_{W_j} \equiv$ West most longitude in j block
$\varphi_{N_j} \equiv$ North most latitude in j block
$\varphi_{S_j} \equiv$ South most latitude in j block.

Expression (5.10) shows that the "Fourier analysis" of the partitioned rows of C_{sz} is immediate, resulting in a "sine" and "cosine" pair per row, twice as many terms as in paragraph 5.1. Furthermore, if the grid has equatorial symmetry, the integrals

$$\int_{\varphi_{S_j}}^{\varphi_{N_j}} \bar{P}_{nm}(\varphi)\cos\varphi\, d\varphi$$

and the resulting $\underline{c}^{nm}_{sz\alpha}$ vectors split naturally into even and odd, as before. Except for the existence of "sine" and "cosine" pairs arising from each column in C^{T}_{sz}, the basic algorithm is applied in the same way as for point data. All important considerations made in the previous example apply to the analysis of area blocks, particularly those pertaining to aliasing.

There remains the question of which way of representing the data, usually not sampled on a regular grid, is least affected by aliasing: whether coefficients estimated from computed area means are necessarily more accurate than those calculated from point values interpolated on a grid. This is not a simple question, though "common sense" (something to be handled with extreme caution) suggests that area means are probably better. This is so because aliasing, as already explained, depends on high frequency terms that are "damped out" by the averaging, while low frequency terms are only slightly modified. In other words, the "signal", or low degree coefficients to be estimated, tends to be greater than the "noise", or higher degree terms that are not estimated.[1] The understanding of the spectral characteristics of data averaged over square blocks is not sufficient, to date, to give a more definite answer. One thing is clear, however: if the coefficients were estimated by collocation from the original ungrided data the resulting mean square error would be the least for any linear estimator utilizing the same data, including those that first "grid" (average) the data and then "collocate" the coefficients from the grided (averaged) data set. This resulting error will include the effect of measurement errors and of aliasing.

5.4 Collocation and Numerical Quadratures

Expression (5.6), in the general case of a nonsymmetric grid, becomes

$$\begin{pmatrix} \hat{C}_{nm} \\ \hat{S}_{nm} \end{pmatrix} = \underline{f}_{\alpha}^{nm\,T}\,\underline{d} = \sum_{i=1}^{N_r} \chi_i^n(m) \begin{pmatrix} c_{nm}^{i\,T} \\ s_{nm}^{i\,T} \end{pmatrix} \underline{d}_i = \sum_{i=1}^{N_r} \sum_{j=0}^{N_c-1} \chi_i^n(m) \begin{pmatrix} \cos \\ \sin \end{pmatrix} \frac{2\pi}{N_c}\, m\, j\, d_{ij} \qquad (5.11)$$

where $\chi^n(m)$ are obtained by solving (3.11). If the data is point data (to simplify the discussion), a common way of computing $\begin{pmatrix} \bar{C}_{nm} \\ \bar{S}_{nm} \end{pmatrix}$ is by "discretizing" the basic expressions

$$\begin{pmatrix} \bar{C}_{nm} \\ \bar{S}_{nm} \end{pmatrix} = \frac{1}{4\pi} \int_{\sigma} \bar{P}_{nm}(\varphi) \begin{pmatrix} \cos \\ \sin \end{pmatrix} m\lambda\; h(\varphi,\lambda)\, d\sigma \qquad (5.12)$$

in the form

$$\begin{pmatrix} \hat{C}_{nm} \\ \hat{S}_{nm} \end{pmatrix} = \frac{1}{4\pi} \sum_{i=1}^{N_r} \sum_{j=0}^{N_c-1} W_{ij}\, \bar{P}_{nm}(\varphi_i) \begin{pmatrix} \cos \\ \sin \end{pmatrix} m\lambda_j\, h(\varphi_i,\lambda_j), \begin{cases} h(\varphi_i,\lambda_j) = d_{ij} \\ \lambda_j = (2\pi/N_c)\, j \end{cases} \qquad (5.13)$$

where the W_{ij} are "quadrature weights", often the area $\Delta\sigma_{ij}$ of each block. Comparing (5.13) to (5.11) and assuming all $P_{nm}(\varphi_i) \neq 0$, one can write:

$$\begin{pmatrix} \hat{C}_{nm} \\ \hat{S}_{nm} \end{pmatrix} = \frac{1}{4\pi} \sum_{i=1}^{N_r} \sum_{j=0}^{N_c-1} W_{ij}^{*}\, \bar{P}_{nm}(\varphi_i) \begin{pmatrix} \cos \\ \sin \end{pmatrix} m\lambda_j\, h(\varphi_i,\lambda_j) \qquad (5.14\text{-a})$$

$$W_{ij}^{*} = \chi_{i,n}^{n}\, \frac{4\pi}{\bar{P}_{nm}(\varphi_i)} \qquad (5.14\text{-b})$$

[1] There is, however, loss of information. If a grid such as that in Figure 2.1-b is used (constant $\Delta\lambda$) then all harmonics of order $m = N_c\, k$, $k = 1, 2, \ldots$, are "averaged out" of the area means.

where the W_{ij}^{*} are the "collocation quadrature weights". As already explained, collocation provides the least aliased coefficients and, for noise data, also the best filtering, of all linear estimation procedures. According to (5.13), in numerical quadratures, regardless of the "weights" used, one computes $\begin{pmatrix} \hat{c}_{nm} \\ \hat{s}_{nm} \end{pmatrix}$ as linear combinations of the data, so estimation by quadratures is linear. Expressions (5.14-a,b) can be used as a justification for considering collocation, in this context, as a numerical quadratures technique with optimal weights for reducing both aliasing and the effect of data errors.

5.5 Estimation of Disturbing Potential from Gravity Anomalies

The value of a function $u(\varphi,\lambda)$ at the "North pole" of a grid, equals the sum of the unnormalized c_{n0} or "zonal" coefficients of its harmonic expansion. For this reason, estimating such value is equivalent to estimating $\sum_{n=0}^{\infty} c_{n0}$, and, when the data is grided: $d(\varphi_i,\lambda_j) = h(\varphi_i,\lambda_j) + n(\varphi_i,\lambda_j)$, this means solving $\underline{\gamma}^{n}(0) = R(0)\,\underline{\chi}^{n}(0)$ for all the N_r "independent"

$$\underline{\gamma}^{n}(0) = [P_{n0}(\sin\varphi_1),\ldots,P_{n0}(\sin\varphi_{N_r})]^{T}$$

and then computing

$$\sum_{n=0}^{\infty} c_{n0} \simeq \sum_{n=0}^{N_r} \hat{c}_{n0} = \sum_{n=0}^{N_r} \sum_{i=1}^{N_r} \chi_i^{0} \sum_{j=0}^{N_c-1} d_{ij}$$

This is a simple explanation of why only the "zero frequency" part of the estimation algorithm has to be carried out.

There is no need, however, of estimating each c_{n0} separately. Assuming that the covariance is isotropic

$$M\{u(P)\,h(Q)\} = \sum_{n=0}^{\infty} c_{uh,n}\,P_n(\psi_{PQ})$$

then, if P is at the pole, the covariance is constant for Q at any position in the same row. The matrix C_{sz}^{T} becomes a $N_r\,N_c$-vector of the type:

$$\underline{C}_{sz}(0) = [\,v_1^{0}\,\underline{c}_0,\ v_2^{0}\,\underline{c}_0,\ldots,\ v_i^{0}\,\underline{c}_0,\ldots,\ v_{N_r}^{0}\,\underline{c}_0]^{T}$$

where $\underline{c}_0 = [1\ 1\ 1\ 1\ \ldots\ 1]^{T}$ is a "zero frequency N_c-vector". Consequently, matrix F^{T} becomes the $N_r\,N_c$-vector

$$\underline{f} = [\,\chi_1^{0}\,\underline{c}_0,\ \chi_2^{0}\,\underline{c}_0,\ldots,\ \chi_i^{0}\,\underline{c}_0,\ldots,\ \chi_{N_r}^{0}\,\underline{c}_0]^{T}$$

where

$$\underline{v}(0) = [\,v_1^{0},\ldots,v_{N_r}^{0}]^{T} \quad \text{and} \quad \underline{\chi}(0) = [\,\chi_1^{0},\ldots,\chi_{N_r}^{0}]^{T}$$

and

$$\underline{\chi}(0) = R(0)^{-1}\,\underline{v}(0)$$

So

$$u(P) = \underline{f}^{T}\underline{d} = \sum_{i=1}^{N_r} \chi_i^{0}\,\underline{c}_0^{T}\,\underline{d}_i = \sum_{i=1}^{N_r} \chi_i^{0} \sum_{j=0}^{N_c-1} d_{ij} = \sum_{i=1}^{N_r} \chi_i^{0} \sum_{j=0}^{N_c-1} (h(Q_{ij}) + n(Q_{ij}))$$

$$= \sum_{i=1}^{N_r} \chi_i^{0}\,\tilde{d}_i \qquad \text{where} \quad \tilde{d}_i = \sum_{j=0}^{N_c-1} d_{ij} \tag{5.15}$$

Expression (5.15) can be interpreted as representing the optimal estimate of $u(P)$ based on the "ring sums" $\sum_{j=0}^{N_c-1} d_{ij}$. It is easy to verify that

$$\underline{\alpha}^T(0) = C_{\bar{d}\bar{d}}(C_{dd} + \tilde{D})^{-1} = \underline{v}^T(0)\, R(0)^{-1}$$

where $\tilde{D}$ is the variance-covariance matrix of $n = \sum_{j=0}^{N_c-1} n_{ij}$ so this interpretation can be used quite consistently. The main consequency of all this is that $u(P)$ can be estimated from the N_r $\bar{d}_i$'s by inverting the $N_r \times N_r$ matrix $R(0)$, instead of the $(N_c N_r) \times (N_c N_r)$ matrix C_{dd}, which can represent great savings in computing, and a corresponding decrease in rounding errors. Setting up $R(0)$, on the other hand, requires just as much effort as forming C_{dd}, because the same number of individual covariances has to be computed for both.

This idea was used by the author as part of his research on the creation of a world height system (see Colombo, 1979), and it will be explained more fully in a future report on that project. The particular application was predicting disturbing potential T from gravity anomalies Δg inside a spherical cap surrounding the point of computation. The semi apertures of the caps studied were 5° and 10°, the rings being spaced at nearly 0.4° intervals. To keep the separation between "gravity stations" roughly constant, the rings were progressively decimated in azimuth towards the center. While this departure from the type of grid considered so far invalidates, in a strict sense, the equivalence between "point" data collocation and "ring averages" collocation, the effect on the results is very small. The covariances were computed using the "two-terms" covariance functions obtained by Jekeli (1978), that have finite recursive form.

Setting up the C_{dd} matrix took near 5 seconds for the 5° cap and 30 seconds for the 10° cap; with 12 "rows" and 475 points in the first case, and 24 "rows" and 3000 points in the second. The times for finding the optimal estimator F (including the inversion of $R(0)$) were of the order of 1 second in both cases. These are C.P.U. times using the Ohio State University's AMDHAL 470V/6-II computer.

6. Conclusions

Minimum variance linear estimation from grided data can be implemented effectively, even when the number of measurements is very large, exploiting the structure induced in the C_{dd} matrix by the regularities of the grid. This is true of both "point data" and of "area means". The restrictions imposed on grids and on covariance functions do not exclude those used in most applications: the "regular" grid and the "isotropic" covariance. The greatest constraint is on the "random" part of the noise, that has to have the same standard deviation at all points in the same "row".

There might be other types of partitions of the sphere that result in even stronger structures for C_{dd}, allowing yet more efficient algorithms for forming and inverting this matrix. Research to this end could be both pleasant and profitable.

The reduced number of computations, when the features of C_{dd} are exploited, means not only shorter computer runs, but also more accurate results.

The algorithm presented here could make possible the computation of spherical harmonic models from global data sets to very high degree and order and with minimum aliasing.

Even after all symmetries in C_{dd} are exploited, creating this matrix is now the major task when doing collocation with very large data sets. This is particularly serious in the case of area means, as the covariances are the "point" covariances integrated twice in two dimensions, as shown in expression (5.9). It would be truly useful to obtain either closed expressions for this covariance, along the lines of Tscherning and Rapp (1976), or else, approximate expressions that are inexpensive to compute.

Storage requirements can be reduced drastically, as only half of the first row of some of the $N_c \times N_c$ partitions C_{ij} of C_{dd} are truly needed; while the $R(k)$ and their inverses are relatively small matrices. The separation of the normal equations into "frequencies" (expression (3.19-b)) makes this approach easy to implement in a parallel-processing machine.

Matrices with the same type of structure considered in this work appear in interpolation and filtering with symmetric kernels of various types, besides covariance functions. In particular, this is true of such things as point mass models when the points are distributed on regular grids.

Already, the ideas presented here have been found useful in estimating disturbing potential from gravity anomalies, because of the economies in computing time they make possible. Soon a program for spherical harmonic analysis based on the same principles will be attempted.

References

Colombo, O. L., A World Vertical Network, Spring Meeting of the American Geophysical Union, Washington, D.C., 1979.

Eren, K., The Spectral Analysis of GEOS-3 Altimeter Data and Frequency Domain Collocation, A Dissertation, Department of Geodetic Science Report, The Ohio State University, Columbus, 1979.

Hobson, E. W., Theory of Spherical and Ellipsoidal Harmonics, Chelsea Editorial, New York, 1965 re-edition of the 1931 version.

Jekeli, C., An Investigation of Two Models for the Degree Variances of Global Covariance Functions, Department of Geodetic Science Report No. 275, The Ohio State University, Columbus, 1978.

Justice, J. H., A Levinson-Type Algorithm for Two-Dimensional Wiener Filtering Using Bivariate Szegö Polynomials, Proc. IEEE, Vol. 65, No. 6, pp. 882-886, New York, 1977.

Kailath, T., Some Results and Insights in Linear Least-Squares Estimation Theory, First Joint IEEE-USSR Workshop on Information Theory, Moscow, 1975.

Katsambalos, K., The Effect of the Smoothing Operator on Potential Coefficient Determinations, Department of Geodetic Science Report No. 287, The Ohio State University, Columbus, 1979.

Kaula, W. M., Theory of Statistical Analysis of Data Distributed Over a Sphere, Reviews of Geophysics, Vol. 5, No. 1, pp. 83-107, 1967.

Lancaster, P., Theory of Matrices, Academic Press, New York, 1969.

Levinson, N., The Wiener R.M.S. (Root Mean Square) Error Criterion in Filter Design and Prediction, J. Math. Phys., Vol. 25, pp. 261-278, 1947.

Luenberger, D. G., Optimization by Vector Space Methods, Wiley & Sons, New York, 1969.

Meissl, P., A Study of Covariance Functions Related to the Earth's Disturbing Potential, Department of Geodetic Science Report No. 151, The Ohio State University, Columbus, 1971.

Moritz, H., Advanced Least Squares Methods, Department of Geodetic Science Report No. 175, The Ohio State University, Columbus, 1972.

Paulik, A., On the Optimal Approximation of Bounded Linear Functionals in Hilbert Spaces with Inner Product Invariant in Rotation or Translation, J. of Computational and Applied Math., Vol. 2, No. 4, pp. 267-272, 1976.

Rapp, R. H., The Relationship Between Mean Anomaly Block Sizes and Spherical Harmonic Representations, J. Geophys. Res., Vol. 82, pp. 5360-5364, 1977.

Rummel, R., A Model Comparison in Least Squares Collocation, Department of Geodetic Science Report No. 238, The Ohio State University, Columbus, 1976.

Rummel, R., and K. P. Schwarz, On the Non-Homogeneity of the Global Covariance Function, Bull. Géodésique, Vol. 51, No. 2, pp. 93-103, Paris 1977.

Sjöberg, L., Potential Coefficient Determinations from 10^{c} Terrestrial Gravity Data by Means of Collocation, Department of Geodetic Science Report No. 274, The Ohio State University, Columbus, 1978.

Sünkel, H., Approximation of Covariance Functions by Non-Positive Functions, The Ohio State University Research Foundation, Project 710334, Report No. 15, Columbus, 1978.

Tikhonov, A. N. and V. Y. Arsenin, Solutions of Ill-Posed Problems, John Wiley & Sons, New York, 1977.

Trench, W., An Algorithm for the Inversion of Finite Hankel Matrices, S.I.A.M. J. Appl. Math., Vol. 13, pp. 1102-1107, 1965.

Tscherning, C. C., and R. H. Rapp, Closed Covariance Expressions for Gravity Anomalies, Geoid Undulations, and Deflections of the Vertical Implied by Anomaly Degree Variance Models, Department of Geodetic Science Report No. 208, The Ohio State University, Columbus, 1974.

Appendix

As explained in Section 2, the matrix of noise covariances D is supposed to consist of Toeplitz circulant blocks, like C_{zz}. If the noise is uncorrelated from measurement to measurement (area mean to area mean) then D is diagonal. To satisfy the Toeplitz condition, this matrix should have those diagonal elements corresponding to points on the same parallel row, at least, equal. In general, this will not be the case. Since approximate results may be better than none, it could be a good idea to use a modified matrix D' instead of D, satisfying all requirements without being "too unlike" D. Two ways of doing this are considered next.

I) $\underline{D' = 0}$

The noise might be ignored altogether, when the quality of the data is very good. Then the problem disappears, because $D \simeq D' = 0$. However, it is desirable to know the effect on the estimates $\hat{\underline{s}}$ of the noise $\underline{n}$

$$\hat{\underline{s}} = F\underline{d} = F(\underline{z} + \underline{n})$$

$$F^T = (C_{zz} + D')^{-1} C_{sz}^T = C_{zz}^{-1} C_{sz}^T \qquad (A.1)$$

F is suboptimal, because the noise matrix has been omitted in (A.1). The variance covariance matrix E of the estimator error is

$$E = M\{(\hat{\underline{s}} - F\underline{d})(\hat{\underline{s}} - F\underline{d})^T\} = C_{ss} - FC_{sz}^T - C_{sz}F + F(C_{zz} + D)F^T$$

(expression (1.2-a)) which can also be written as

$$E = E_a + E_n = [C_{ss} - FC_{sz}^T - C_{sz}F + FC_{zz}F^T] + [FDF^T] \qquad (A.2)$$

Here E_a represents the error due to the way the data is sampled (aliasing), and E_n is the contribution of the "propagated" noise. Forming E, E_a, and E_n requires calculating a number of matrix products, the most involved of which is $FC_{zz}F^T$. Another matrix product of interest is $C_{zz}F^T$, needed for the second part of this Appendix. These products can be obtained efficiently by exploiting the structure of C_{zz}, which is the same as that of C_{zz}^{-1}. Therefore, the method for solving the system of equations $\underline{y} = C_{zz}\underline{x}$ (i.e., finding $C_{zz}^{-1}\underline{y}$) can be applied to obtaining $C_{zz}\underline{y}$ as well, with minor modifications. To this end, the algorithm described in paragraph (3.3) has to be changed as follows:

a) Use the matrices R(k) and the vectors $\underline{x}_{\alpha,n}^{k} = [\lambda_{\alpha,1}^{k}(\frac{C_k}{\underline{s}_k}), \ldots, \lambda_{\alpha,N_k}^{k}(\frac{C_k}{\underline{s}_k})]^T$ (where $\underline{f}^n = \sum_{k=0}^{N}\sum_{\alpha=0}^{1} \underline{x}_\alpha$ is the n th column of F^T), which have been obtained during the determination of F, to calculate

$$\underline{\xi}_\alpha^n(k) = R(k)\,\underline{x}_\alpha^n(k) \qquad (A.3)$$

where $\underline{\chi}^{n}_{\alpha}(k) = [\chi^{k}_{\alpha,1,n}, \ldots, \chi^{k}_{\alpha,N_r,n}]^T$

b) Form the columns of $H = C_{zz} F^T$ by "Fourier synthesis":

$$\underline{h}^n = C_{zz}\underline{f}^n = \sum_{k=0}^{N} \sum_{\alpha=0}^{1} \underline{e}^{k}_{\alpha,n} \tag{A.4}$$

where $\underline{e}^{k}_{\alpha,n} = [\xi^{k}_{\alpha,1,n}\left(\frac{c_k}{s_k}\right), \ldots, \xi^{k}_{\alpha,N_r,n}\left(\frac{c_k}{s_k}\right)]^T$

b') In particular, the elements of $J = F C_{zz} F^T$ are

$$\begin{aligned} j_{ij} = \underline{f}^{iT} C_{zz} \underline{f}^{j} = \underline{f}^{iT} \underline{h}_j &= \sum_{k=0}^{N} \sum_{\alpha=0}^{1} \underline{\chi}^{kT}_{\alpha,i} \sum_{\beta=0}^{N} \sum_{\beta=0}^{1} \underline{e}^{*}_{\beta,j} \\ &= H_k \sum_{k=0}^{N} \sum_{r=1}^{N_r} \sum_{\alpha=0}^{1} \chi^{k}_{\alpha,r,i} \quad \xi^{k}_{\alpha,r,j} \end{aligned} \tag{A.5}$$

where H has been defined in paragraph (3.2).

II) Replacing D with an "Average Noise" Matrix, and Refining

If the noise is white, but its standard deviation varies from point to point (or block to block) along a parallel, a diagonal D' matrix could be chosen with all elements corresponding to the n th data row equal to the average of the variances of the individual data in that row:

$$d'_{nn} = \bar{\sigma}_n = \frac{1}{N_c} \sum_{i=0}^{N_c-1} \sigma_{ni} \tag{A.6}$$

This could be regarded as a reasonable approach in cases where the noise is too great to be neglected without serious loss of accuracy. Whichever way the estimator F is obtained, it can always be "refined" (if suboptimal) to make it more accurate. One way of doing this is to apply "steepest descent", where the variances of the errors (diagonal elements of E) are reduced simultaneously along the "line" of search defined by the present F and the matrix gradient of the trace of E, $\frac{\partial t_r}{\partial F}(E)$, to obtain the "improved" estimator F_s:

$$\begin{aligned} F_s^T &= F^T + [C_{dd} F^T - C_{sz}^T] K \\ &= F^T + [C_{zz} F^T + D F^T - C_{sz}^T] K \end{aligned} \tag{A.7}$$

K is diagonal, and

$$k_{nn} = \frac{-\nabla_n^T \nabla_n}{\nabla_n^T C_{dd} \nabla_n}$$

$$\nabla_n = C_{dd}\underline{f}^n - \underline{c}^n_{sz} = C_{zz}\underline{f}^n + D\underline{f}^n - \underline{c}^n_{sz}$$

(where $\underline{f}^n$ and $\underline{c}^n_{sz}$ are the nth columns of F^T and C_{sz}^T, respectively). The element k_{nn} constitutes the optimal "step length" for $\underline{f}^n$. Expression (A.7) requires calculating $C_{zz} F^T$ and $\underline{f}^{nT} C_{zz} \underline{f}^n$. This can be done using the procedure explained in the first part of this Appendix, if those products are not already available. The whole procedure can be iterated.